



Big Data Analytics in The Cloud

Ekasit Kijsipongse

Researcher

ekasit.kijsipongse@nectec.or.th, kitwai@nectec.or.th

Large-Scale Simulation Research Laboratory (LSR)
Data Analytics and Computing Research Unit (DACRU)
NECTEC

Session 5 “HPC meets Big Data”

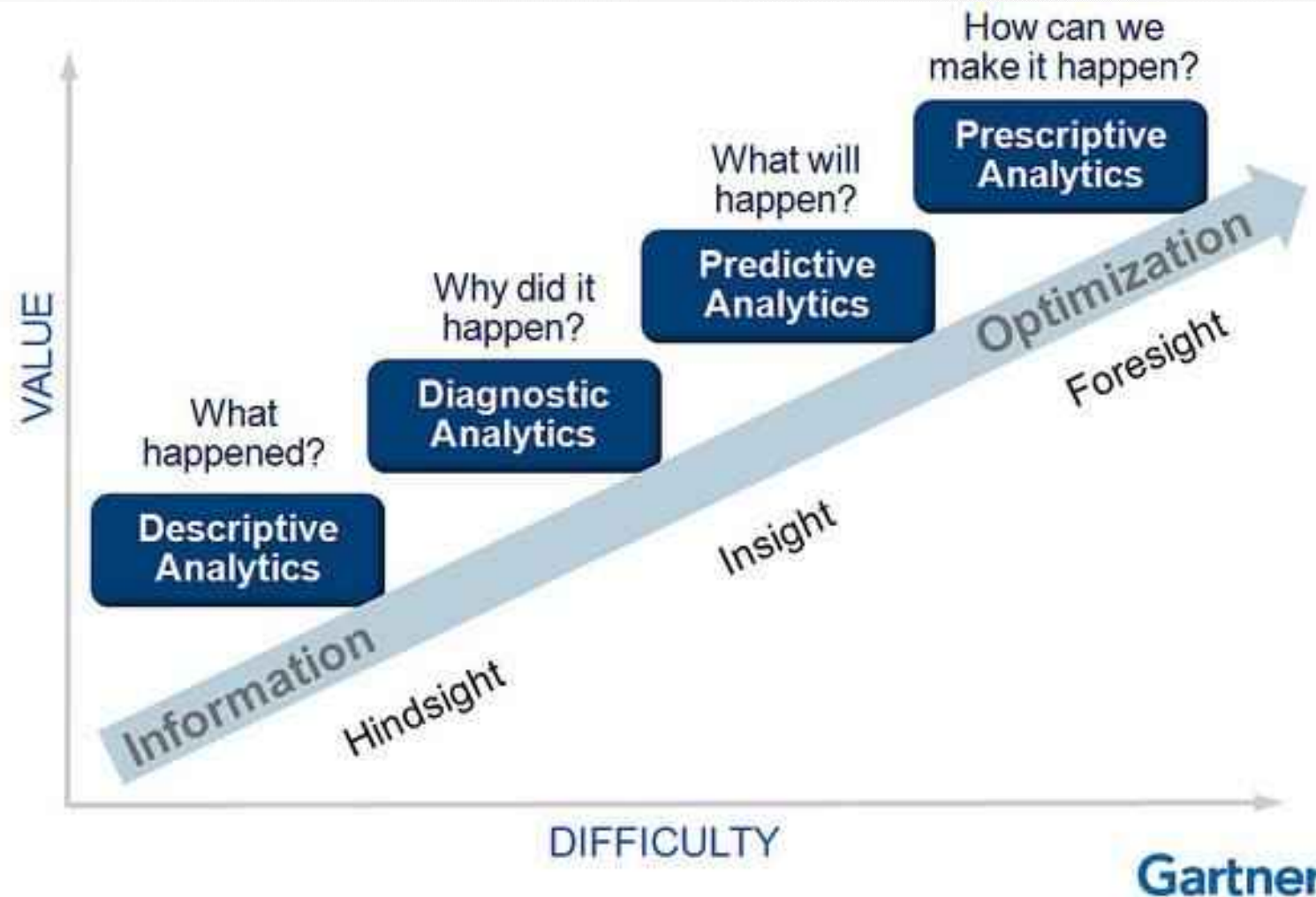
วันจันทร์ที่ 11 กันยายน 2560 | Meeting room 2 ศูนย์ประชุมแห่งชาติสิริกิติ์

Data-Driven Organization

“Data is a new oil”



Types of Data Analytics



Big Data and HPC Applications



Sales Forecast



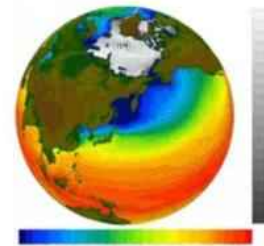
Social Network



Bank



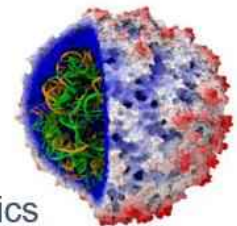
Retail Store



climate & weather
prediction simulations

CATACACGT
GCATGCATA

Genomic sequencing



Molecular Dynamics



manufacturing & material
science



Oil and Gas

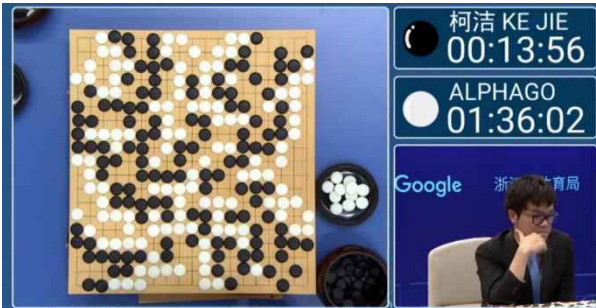


New Applications Need More Data

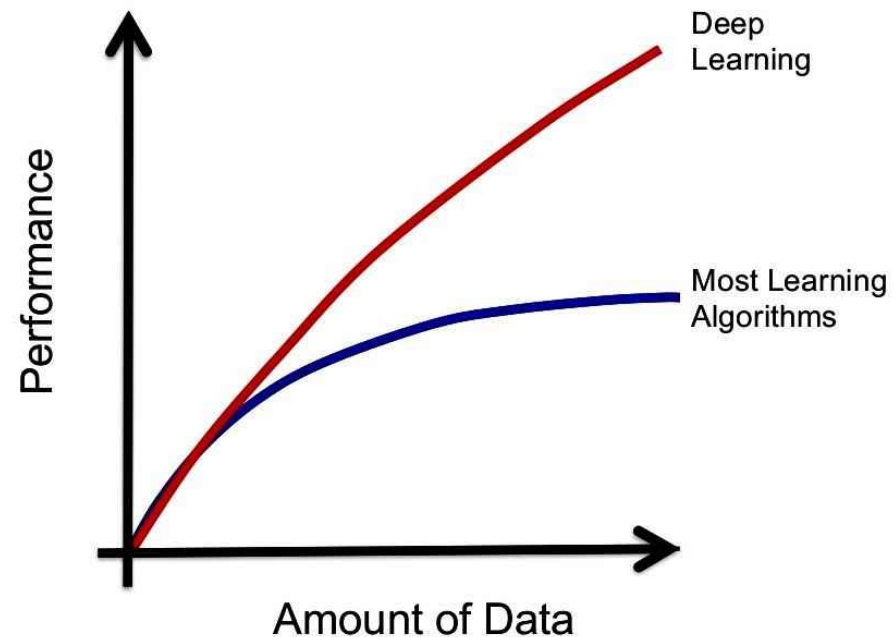
Self-driving cars



By Steve Jurvetson [CC BY 2.0]



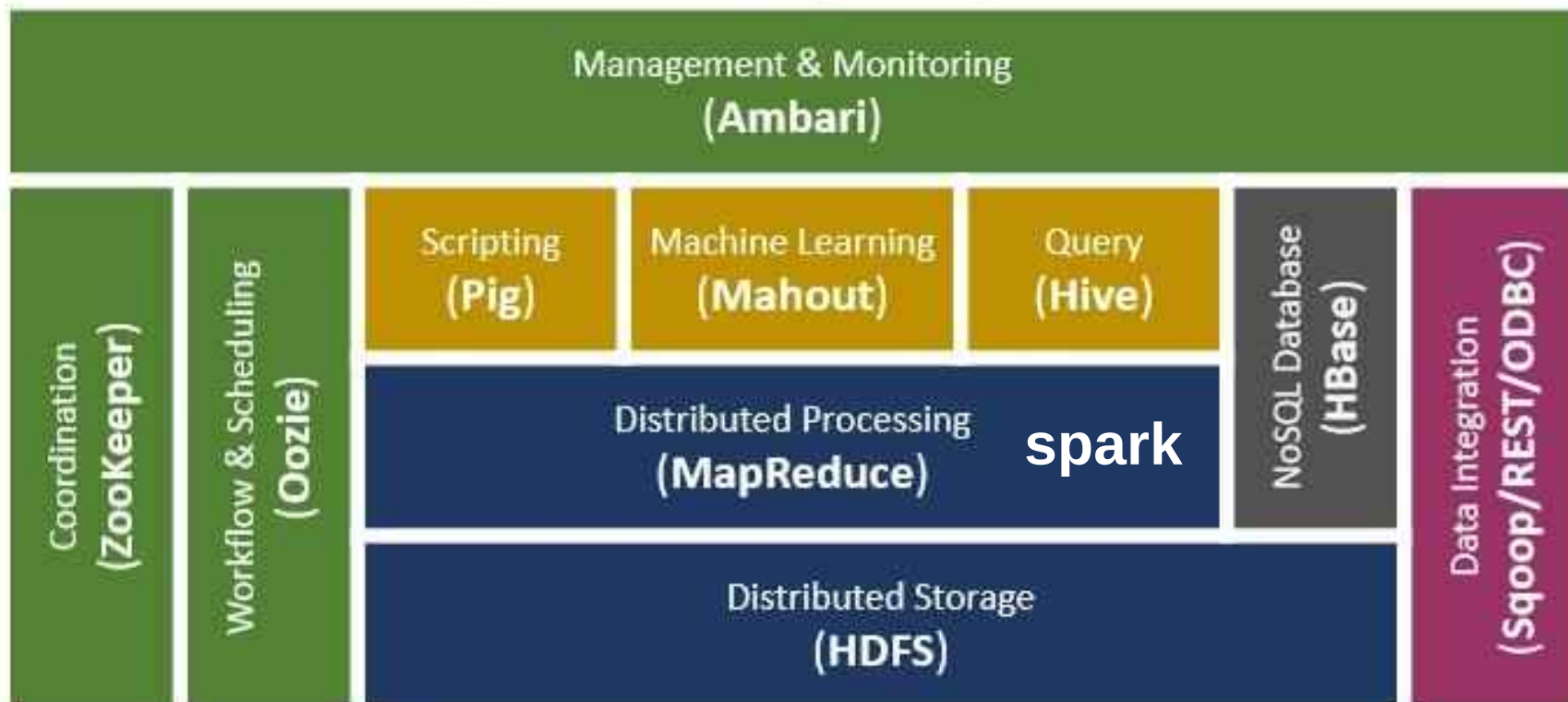
BIG DATA & DEEP LEARNING



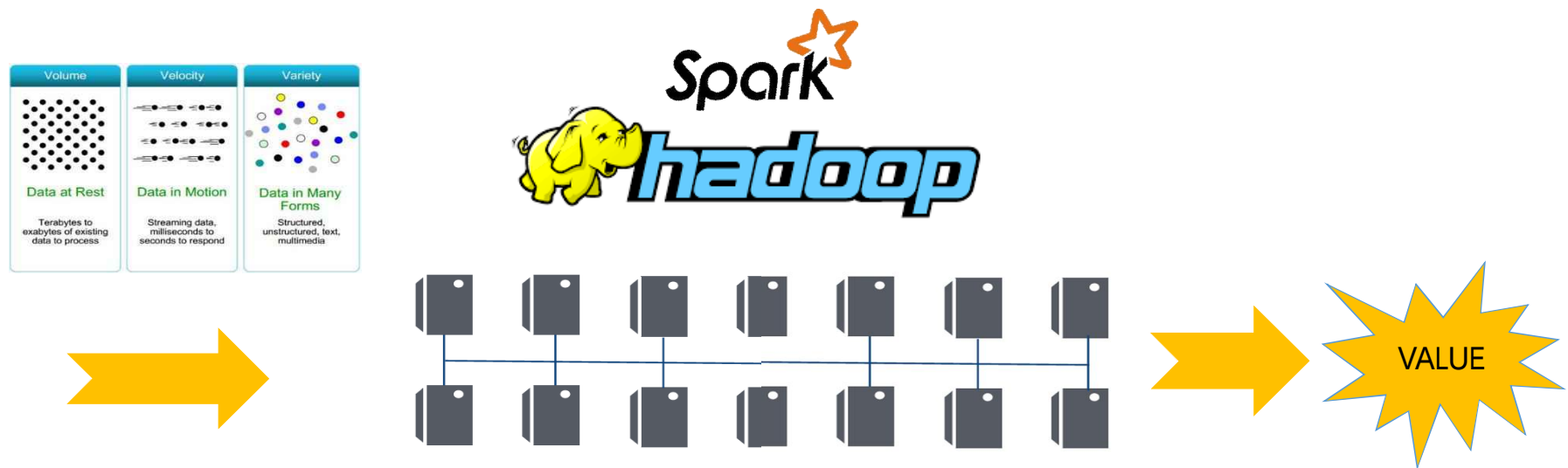
<https://www.linkedin.com/pulse/deep-learning-aviation-francis-j-duque>

Big Data (Hadoop) Architecture

Apache Hadoop Ecosystem



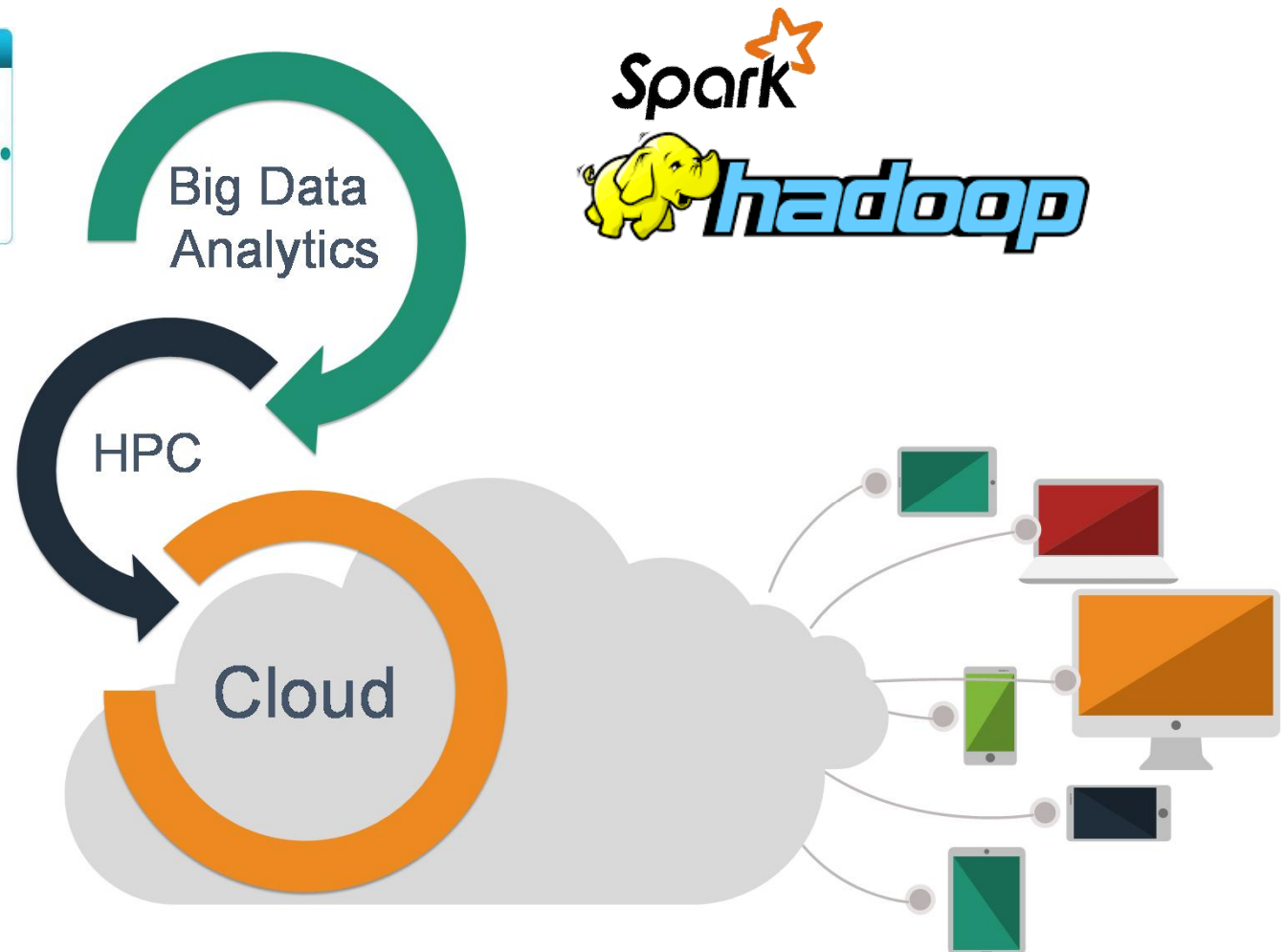
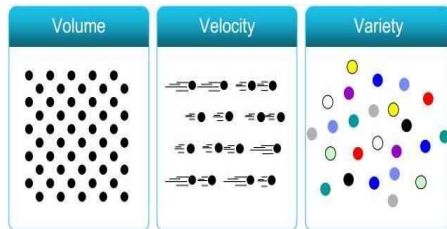
Distributed and Parallel Computing



Big Data Analytics Meets HPC



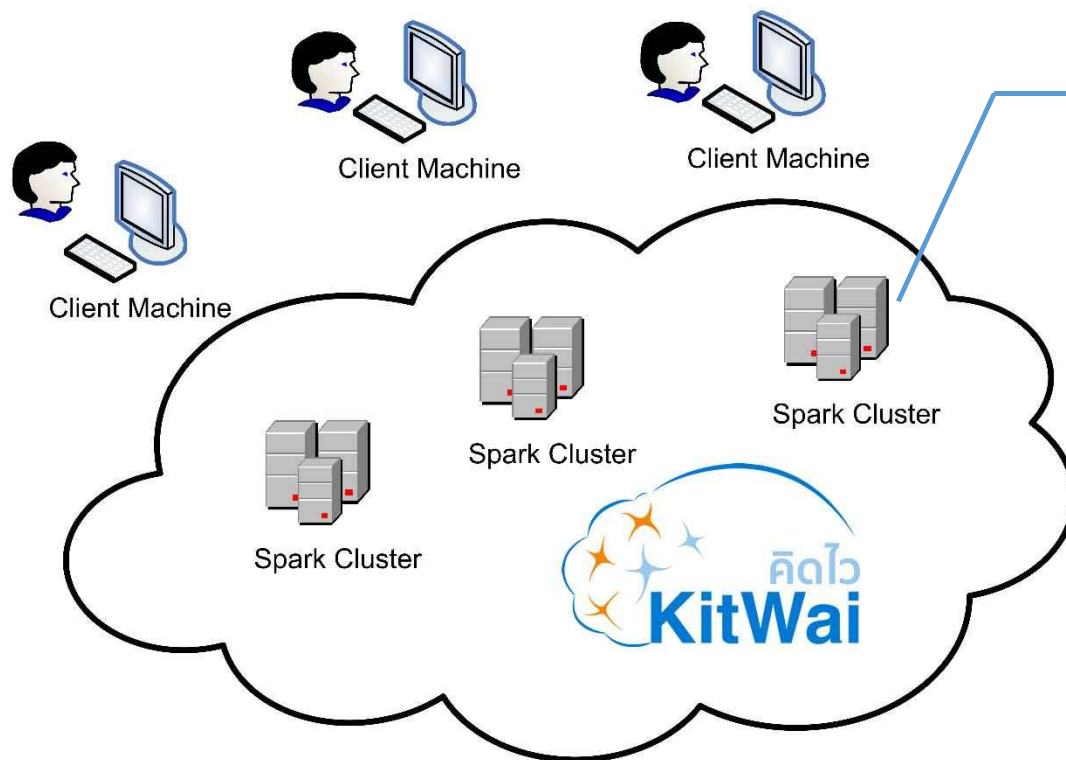
Big Data Analytics as a Service



Hadoop in The Cloud



KitWai Cloud Platform for Data Analytics



Resource monitoring
Web-based notebook
Web-based data analytics
GPU enabling
JDBC/ODBC connection

-  On demand self-service
-  Hadoop/Spark Cluster
-  OpenStack Cloud

HPC Hardware for Big Data Analytics

Data intensive Computing

- ✓ Large memory and SSD
- ✓ GPU with fast CPU-GPU and GPU-GPU connection (Nvidia NVLink)
- ✓ Infiniband

NSTDA HPC/DA Platform (2018)



Support Research from NECTEC, BIOTEC, MTEC, and NANOTEC

Cloud Platform for Data Analytics (DA)

การนำข้อมูลมาวิเคราะห์ในเชิงลึก เพื่อ
แก้ปัญหา หรือให้เกิดประโยชน์ในเรื่อง
หนึ่งๆ สนับสนุนงานในรูปแบบ Data
Driven and Statistical Analysis เช่น

- Genome Database Service
- Data Analytics Platform
as a Service
- AI Research : Deep Learning,
Language Processing, Behavior
Analysis

High Performance Computing (HPC)

การประมวลผลด้วยคอมพิวเตอร์
สมรรถนะสูง สนับสนุนงานในรูปแบบ
Computational Model and
Scientific Computing เช่น

- การประมวลผลข้อมูลจีโนม
ของประเทศ
- Computational Design Studio
- Computational Fluid Dynamics
- Quantum Chemistry

NSTDA HPC/DA Platform (2018)



140 Compute nodes, 5000 cores, 30 TB RAM, 36 Nvidia GPUs, 425 TB Storage, 320 TB Backup

